ANALYSIS OF CHALLENGES, OPEN RESEARCH ISSUES AND TOOLS BASED ON SURVEY OF BIG DATA ANALYTICS

V.T.KRUTHIKA
Research Scholar,
Department of Computer Science,
Adaikalamatha College,
(Affiliated to Bharathidasan University)
Vallam, Thanjavur, Tamil Nadu, India.

Dr.L.NAGARAJAN
Research Advisor,
Director,
Department of Computer Science,
Adaikalamatha College,
(Affiliated to Bharathidasan University)
Vallam, Thanjavur, Tamil Nadu, India.

ABSTRACT

A vast terabyte database is built using existing data systems, digitally-equipped technologies and the Internet of Things. This enormous data analysis needs considerable effort to extract decision-making information at several levels. Big data searches have gained significant attention in recent years as localization applications have increased exponentially. Current Big Data systems, however, are disc based and do not meet superior efficiency and short reaction times. As data in distributed memory environments increases, consumers desire data processing to be modest. The additional benefit is that it presents a new perspective for academics to go into open-ended research issues. This idea provides a way by which big data analysis may be carried out within SOMA (Scalable and Operational Memory Analytics) environment. SOMA features a Data Analysis Index System of two levels.

INDEX TERMS: Data Storage, Visualization, Quantum Computing.

INTRODUCTION

During the digital era, data is generated from various sources, resulting in big data sets. In several disciplines, it offers evolutionary advances with big data sets overall it refers to massively complex datasets that are not easily processed by predicting typical database management tool use or application development with the use of databases. These files are accessible in organized, semi-structured and unstructured versions in petabytes and beyond. That range

of voltages is formally described as 3V to 4V. Three Vs are used to describe volume, speed, and version. Volume is the abundance of information produced on an everyday basis, while speed is increasing and the data is gathered for analysis as fast as possible. Various enormous growth dates were created of mobile phones and location-based services, usually a collection of geo-tagged text segments (LBS). A keyword space query is often a geographic prediction and sophisticated keyword predicates aimed at locating places with the same text descriptions in the area.

LITERATURE SURVEY

BIG DATA ANALYTICS CHALLENGES

Recently, several areas in which large amounts of data exist, such as healthcare, government, retail, biochemistry, and many more interdisciplinary research studies, have been collecting their data. Taking these features into consideration, Big Data provides future researchers different perspectives regarding for the processing of knowledge. Oppositions nevertheless, are always challenging.
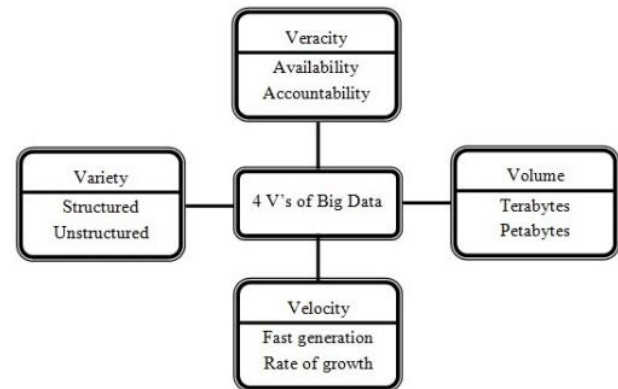
1. DATA STORAGE AND ANALYSIS

Over the last many years, several new ways to store data have emerged, including mobile devices, air sensors, RFI, and other kinds of sensors. These data are saved on expenses, although they were taken into account, despite or erased since insufficient capacity to retain them. At the beginning of the study, large data analysis is limited by storage medium and greater input/output rates. Another major difficulty in the research of quantity of big data. Data mining duties have considerably risen with the ever-growing of datasets. Here huge information analysis involves more computational complications.

The major difficulty is the management of incomprehensibility and uncertainty in the database. In common, the complexity of the computer is systematically modelled. It might be difficult to construct a complete mathematical framework, widely utilized for huge data. But in particular problems of a domain data analysis may be easily understood. Data reduction, data selection and method selection are key responsibility, mainly in the huge datasets.

2. COMPUTATIONAL COMPLEXITIES AND KNOWLEDGE DISCOVERY

The difficult search and display of data is due to the sheer amount of data. The following subcategories are featured in the digital preservation subsection: authentication, archiving, administration, preservation, informatics, and reproduction. Large data set analysis involves more computational complications.

An extensive mathematical framework that can be utilized in large quantities might be difficult to construct. However, specifically, the particular problems of a domain data analysis are easy to understand. But the present techniques for big-data analysis lack computer complexity, ambiguity and inconsistency in their treatment, and have low performance in IJACSA. The growth of technology and chances which can address computer flexibility, uncertainty and inconsistency effectively brings us a major difficulty.

3. RAPID DATA SCALING AND VISUALIZATION

The major issue is the study of Big Data technology is scalability and safety. In recent decades, researchers have been following Moore's law to speed up data analytics and computers. Sampling, on-line, and multiple analytic perspective which are essential for the former. Incremental methods brings us a strong scalability in the examination of huge data. Since data size is quicker than CPUs, the process of technology is naturally changing dramatically expanded by the number of nodes. Big data producing numerous

challenges for both hardware and software advances towards to the similar computing, cloud computing, dis-computer, visualization, scalability. It may be seen from large numbers. To address this new challenge, we will need to combine more mathematics with computer science. In order to address this new challenge, we will have to combine more mathematics with computer science.

THE INFORMATION SECURITY FIELD

Massive data volumes are correlated, examined and explored for major patterns in large data analyses. There are different policies for every organizations to protect sensitive information. In analysis of big data the preservation of important information is a key problem. Big Data poses an enormous security concern. Consequently, the information security is becoming a major challenge of analysis. Big data reliability may be improved by employing authentication, permission and encryption approaches. There are several types of security approaches, one of which is the use of big data. This big data collection approach includes things like network size, various devices, Real-time safety surveillance, and the absence of an infiltration mechanism.

BIG DATA ANALYTICS OPEN RESEARCH ISSUES

Organizations or institutions devote a lot of attention to analytics of big data and data science. It involves digging out enormous amounts of data and information. Applying novel modelling, machine learning, uncertain

data analysis, sample identification, storage of data, statistical learning and signal processing in the context of data science and big data are included under a category of big data and data science applications. When appropriate technology integration and analysis are employed, the following set of occurrences may be predicted. Subjects that do not have to do with closed topics in Big Data Analytics are covered in this section. There are three major research areas to deal with in order to do analysis of big data: the cloud computing, internet of things (IoT) and BIC.

1. BIG DATA ANALYTICS OF IOT

In worldwide internet connections, business art, cultural revolutions, and some very big personalities have been reassembled. Computers are now working on establishing control over thousands of different pieces of equipment and developing the Internet of Things (IoT). So, like people with web browsers, devices become the user of the Internet. The collection of knowledge from IoT Data for big data experts is the most crucial. In order to value IoT data, it is therefore it is required to create infrastructure.

2. BIG DATA ANALYTICS ON CLOUD COMPUTING

The new virtualization technologies made supercomputing accessible and cheaper. Computing infrastructures covered by virtualization software, generate systems like a real computer, however, free to detail requirements such as multiple processes, disc space, storage and operating structure. It will mean cloud computing and big data getting a tremendous of improvement out of anything that stands in the way.
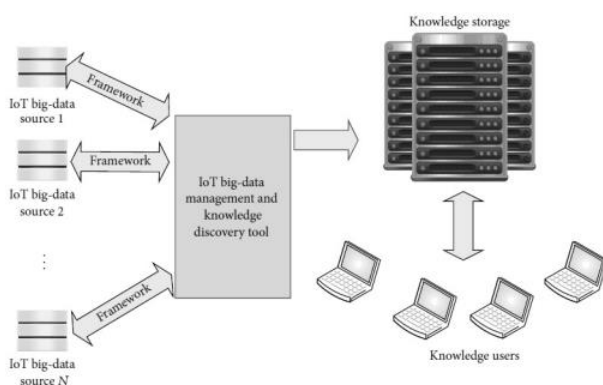
3. BIG DATA ANALYTICS ON BIO-INSPIRED COMPUTING

On the contrary, computing with bio-inspiration was not adopted as a solution to any Problems in the real-world. Without middle control, genetic structured are organized. Search and locate the finest data service solution to reflect information management costs and service maintenance for a bio-inspired mechanism. These approaches are provided by biological molecules like DNA and protein for the calculation of information collection, memory and processing.

4. QUANTUM COMPUTING FOR BIG DATA ANALYSIS

Quantum computers can handle a large amount of data at the same time, even if they are physically small. It might be feasible to enhance this exponentially with computer systems. If a real quantum computer is now available, problems that are, of course, extremely difficult to handle today's enormous data problems may also be

addressed. Although the technological difficulty in constructing a quantum computer may be increasing, the most important one may soon be solved. A mechanism for combining quantum mechanics for information treatment is provided by quantum computing.

BIG DATA PROCESSING TOOLS

There are several tools that may be used for processing big data. This section covers several current approaches to Big Data Analytics, with a particular focus on three important new technologies: Map Reduce, Apache Spark, and Storm. Most technologies aim to be used in batch processing, and they also have the features that are best for streaming and interactive analysis. More used than any other tool in the Hadoop architecture is Mahout and Dryad for batch processing.
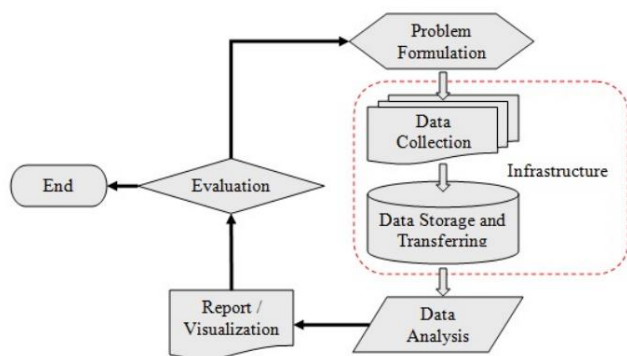
1. TECHNIQUES OF MAP REDUCE

It's clear that the most established systems for Big Data analysis are Hadoop and MapReduce. The described programmers are: Hadoop Distributed File System (HDFS), MapReduce Apache Hive, Hadoop and.

Reduction the size of a map can be referred to as the practice of dividing and conquering. In this scenario, splitting and winning is implemented in two parts, such as the map and decreasing stages.

2. BIG DATA PROCESSING

Many tools for processing large data may be found. In this part, we examine a numerous of contemporary big data analytics approaches, towards on three major new technologies Apache Spark, Map Reduce, and Storm. Batch processing, pro-processing streaming, and interactive analysis are the technologies provided in the majors. Batch processing is now done using the Hadoop architecture, namely with tools like Mahout and Dryad.

In Spark, RDD (Resilient Distributed Datasets) are vital, as the data is stored in memory with no replication. It facilitates repeated calculation, enhances speed and the usage of resources. The major advantages is that it supports streaming data as well as Map Reduce, machine learning techniques and graph algorithms.

CONCLUSION

It has become more dramatic how much data there is in the world nowadays. For a guy without a scientific mind, it's tough to go through the facts. We work to accomplish this goal by reviewing all of the many research topics, issues, and tools applied to measuring this huge dataset. According to this study, each big data platform is said to focus on just one application. We anticipate that the

techniques outlined here will be given greater consideration by future academics to tackle difficult data challenges. Some work well in a setting of batch processing, while other peoples are better for real-time work. The size of each platform with large amounts of data will offer distinct capabilities. Statistics, machine learning, data extraction, data stream processing, cloud computing, quantum computing, and data analysis all have many analytical approaches used in them.

REFERENCES

[1] A survey of data partitioning and sampling methods to support big data analysis: Mohammad Sultan Mahmud; Joshua Zhexue Huang; Salman Salloum; Tamer Z. Emara; Kuanishbay Sadatdiynov, Big Data Mining and Analytics_2020.

[2] Appling big data based deep learning system to intrusion detection: Wei Zhong; Ning Yu; Chunyu Ai, Big Data Mining and Analytics_2020.

[3] A novel clustering technique for efficient clustering of big data in Hadoop Ecosystem: Sunil Kumar; Maninder Singh, Big Data Mining and Analytics_2019.

[4] Big Educational Data & Analytics: Survey, Architecture and Challenges: Kenneth Li-Minn Ang; Feng Lu Ge; Kah Phooi Seng, IEEE Access_2020.

[5] Error data analytics on RSS range-based localization: Shuhui Yang; Zimu Yuan; Wei Li, Big Data Mining and Analytics_2020.

[6] Model-Based Big Data Analytics-as-a-Service: Take Big Data to the Next Level: Claudio Agostino Ardagna; Valerio Bellandi; Michele Bezzi; Paolo Ceravolo; Ernesto Damiani; Cedric Hebert, IEEE Transactions on Services Computing_2021.

[7] A Novel Data-Driven Situation Awareness Approach for Future Grids—Using Large Random Matrices for Big Data Modeling: Xing He; Lei Chu; Robert Caiming Qiu; Qian Ai; Zenan Ling, IEEE Access_2018.

[8] Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management: Dinithi Nallaperuma; Rashmika Nawaratne; Tharindu Bandaragoda; Achini Adikari; Su Nguyen; Thimal Kempitiya; Daswin De Silva; Damminda Alahakoon; Dakshan Pothuhera, IEEE Transactions on Intelligent Transportation Systems_2019.

[9] A Methodology of Real-Time Data Fusion for Localized Big Data Analytics: Sohail Jabbar; Kaleem R. Malik; Mudassar Ahmad; Omar Aldabbas; Muhammad Asif; Shehzad Khalid; Kijun Han; Syed Hassan Ahmed, IEEE Access_2018.

[10] Hybrid recommender system for tourism based on big data and AI: A conceptual framework: Khalid Al Fararni; Fouad Nafis; Badraddine Aghoutane; Ali Yahyaouy; Jamal Riffi; Abdelouahed Sabri, Big Data Mining and Analytics_2021.

[11] Deep Learning for IoT Big Data and Streaming Analytics: A Survey: Mehdi Mohammadi; Ala Al-Fuqaha; Sameh Sorour; Mohsen Guizani, IEEE Communications Surveys & Tutorials_2018.

[12] Cost-Effective Cloud Server Provisioning for Predictable Performance of Big Data Analytics: Fei Xu; Haoyue Zheng; Huan Jiang; Wujie Shao; Haikun Liu; Zhi Zhou, IEEE Transactions on Parallel and Distributed Systems_2019.

[13] An Integrated Big and Fast Data Analytics Platform for Smart Urban Transportation Management: Sandro Fiore; Donatello Elia; Carlos Eduardo Pires; Demetrio Gomes Mestre; Cinzia Cappiello; Monica Vitali; Nazareno Andrade; Tarciso Braz; Daniele Lezzi; Regina Moraes; Tania Basso; Nádia P. Kozievitch; Keiko Verônica Ono Fonseca; Nuno Antunes; Marco Vieira; Cosimo Palazzo; Ignacio Blanquer; Wagner Meira; Giovanni Aloisio, IEEE Access_ 2019.